Metagenomic Discovery and Screening of Novel Recombinase Proteins for Targeted Integration

Joseph Neumeyer, Meng Wu, Shreya Jambardi, Amy Zhao, Bradley Taylor, Georgia Giannoukos, Eugenio Marco, Hui Li

Editas Medicine, Inc., Cambridge, MA, USA.

editas MEDICINE

Poster 1681

OBJECTIVE

- The goal of this study was to discover large serine recombinases (LSRs) that are functional in human cells as a foundation for a novel gene editing technology that can integrate large transgenes into the human genome efficiently *in vivo*.

INTRODUCTION

- Current technologies to integrate multi-kilobase (kb) DNA sequences *in vivo*, such as retroviruses, transposases, or others mediated by homology directed repair (HDR), have limitations.
- Site-specific large serine recombinases (LSRs) are used by phages to integrate the phage genome into a bacterial genome without relying on endogenous host DNA repair machinery. They also have the potential to integrate into specific target sites.
- Through metagenomic data mining using the LSR-finder, a bioinformatics pipeline searching algorithm developed by Editas, we discovered thousands of potential LSR candidates and reconstructed their cognate DNA recognition sites (attB/attP).
- Using high-throughput functional screening, we discovered hundreds of highly potent LSRs that can integrate into the human genome with various specificities and efficiencies.

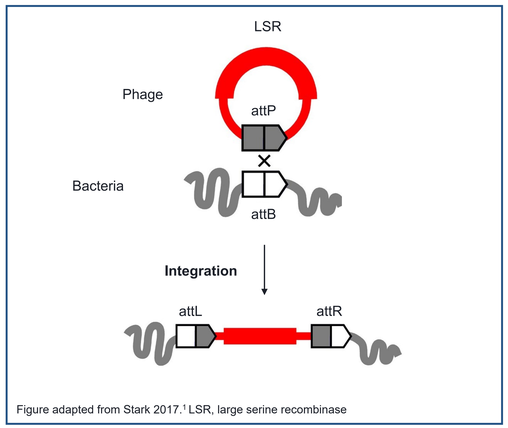
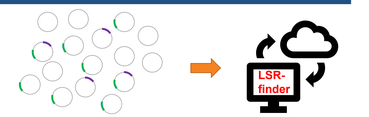
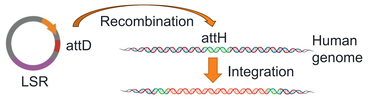


Figure adapted from Stark 2017.[1] LSR, large serine recombinase

METHODS

- LSRs were discovered metagenomically from 918,829 bacterial genomes in public databases.
- Functional LSRs were identified via individual and pooled screening in human cells.

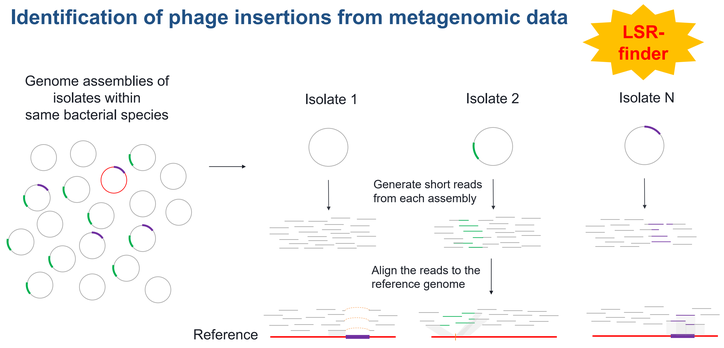




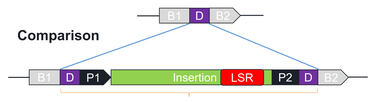
attD, donor recognition site; attH, human genome recognition site; LSR, large serine recombinase.

RESULTS: Computational Discovery of LSRs in Bacterial Genomes

Identification of phage insertions from metagenomic data



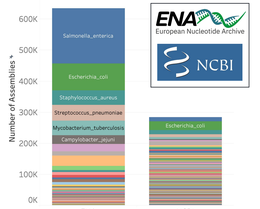
Annotation of LSR proteins and corresponding attB/attP sites



- Annotate LSR protein with all three domains: resolvase, recombinase (RD), and zinc ribbon (ZD) domains
  - LSR (400–600 AAs)
- Reconstruct attB and attP from the sequences surrounding the phage insertion boundary
  - attB = B1 + D + B2 (~40 bps)
  - attP = P2 + D + P1 (~50 bps)

AA, amino acid; B, bacteria; D, central dinucleotide; LSR, large serine recombinase; P, phage; RD, recombinase domain; ZD, zinc ribbon domain.

Discovery of novel LSRs



635,246 (ENA) + 283,583 (NCBI) bacterial genomes

↓ Identify phage insertions

34,636 phage insertions

↓ Have all resolvase, recombinase, and zinc ribbon domains

25,149 LSR candidates

↓ Remove redundant LSRs by protein and attB/P sequence

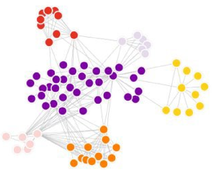
12,713 putative LSRs + attB/attP

ENA, European Nucleotide Archive; LSR, large serine recombinase; NCBI, National Center for Biotechnology.

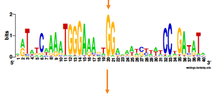
Cluster LSRs to select screening candidates

Cluster LSRs by their three protein domains (>90% identity):

- 159 large clusters with at least 10 LSRs
- ~6,300 LSRs are included in these 159 large clusters

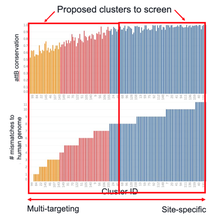


Generate conservation score based on attB diversity



Select a representative LSR from each of the 159 large clusters to screen

Generate two pooled libraries based on the attB conservation score (cutoff = 0.9)

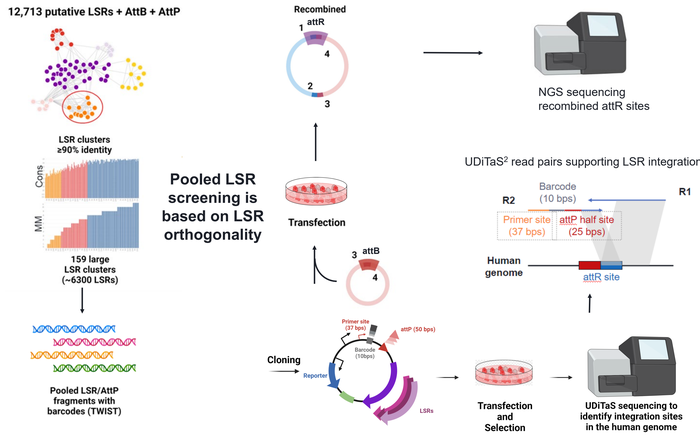


LSR, large serine recombinase.

RESULTS: Platform for Screening Functional Recombination of Putative LSRs

Pooled screening of orthogonal LSRs in human cells

Pooled plasmid recombination assay



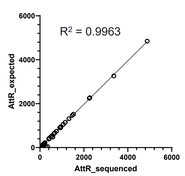
Pooled human genome integration assay

The plasmid recombination assay determines the LSR recombination activity in human cells, while the genome integration assay investigates the efficiency and specificity of LSR integration into the human genome.

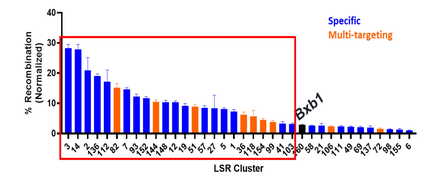
LSR, large serine recombinase; NGS, next generation sequencing; bps, base pairs; UDiTaS, Uni-Directional Targeted Sequencing methodology.

Highly active and specific LSRs are identified to work in human cells

Pooled plasmid recombination assay

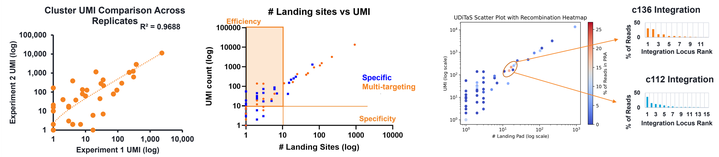


We demonstrated the accuracy of LSR attB/attP *in silico* prediction and LSR orthogonality with the potential of multiplexed integration



Multiple LSR clusters demonstrated higher recombination activity than *Bxb1*, a well-characterized LSR

Pooled human genomic integration assay



Many LSRs show genomic integration across two experiments

LSRs with a high number of UMIs and a low number of landing sites have potential as gene editing medicines

Plasmid recombination assay data overlaid with UDiTaS results shows LSRs with efficient recombination and genomic integration

Multiple LSR clusters were identified to have significant efficiency and specificity of integration into the human genome and have the potential as a new platform for gene editing.

LSR, large serine recombinase; UDiTaS, Uni-Directional Targeted Sequencing methodology; UMI, unique molecular identifiers.

CONCLUSIONS

- The LSR-finder, a bioinformatics pipeline searching algorithm developed by Editas, can discover functional LSRs with a high degree of accuracy.
- Several thousand candidate LSRs with their attB/attP sequences were identified from public metagenomic databases using the LSR-finder.
- 159 representative LSRs were selected based on a clustering algorithm to represent the majority of LSRs for high-throughput functional screening in human cells.
- Hundreds of novel LSRs showed potent recombination and genomic integration activity and specificity in human cells.
- These recombinase proteins may allow for the development of novel gene editing technologies capable of knocking in large transgenes *in vivo*, potentially enabling the targeting of additional therapeutic indications.

REFERENCES

1. Stark WM. *Curr Opin Microbio* 2017; 38: 130–136.
2. Giannoukos G *et al. BMC Genomics* 2018; 19 (1): 212.

DISCLOSURES

All authors are current or former employees and shareholders of Editas Medicine, Inc.

The authors have filed a patent application on the data presented here.

Acknowledgments:
Pei Lin, Shailesh Gurung, Oscar Velazquez Cuazitl, Sean Scott, Deric Zhang, Mihnea Bulugioiu, Elise Thompson, Kayla Delano, James Bochicchio, Lily Maxham, Deric Zhang, Michael Dinsmore, John Zuris, Christopher Wilson, Adrian Timmers, Donato Aceto, Scott Douglas, and Andrew Davenport.